SimpliFiTM in a Box: Multiomics Data Analysis on Private Servers

Jim Palmeri¹, Darryl J.C. Pappin²; John Wilson¹

¹ProtiFi LLC, Fairport, NY, United States; ²Cold Spring Harbor Laboratory, Long Island, United States



HE RAPID EXPANSION OF OMICS DATA poses a significant challenge in understanding and extracting actionable insights. This challenge is compounded by the increasing sample sizes and diversity of omics analyses within studies.

O ADDRESS THIS CHALLENGE, speed the path to understanding and open the power of omics to non-experts, we developed SimpliFi[™], an interactive, intuitive data-to-meaning engine. Employing nonparametric statistics based on inherent data structure, SimpliFi[™] utilizes resampling techniques to calculate confidence intervals for all metrics, including p-values. This approach highlights often-overlooked uncertainties in statistical metrics. SimpliFi[™] handles all kinds of omics data and seamlessly integrates data from various omics analyses. Sharing, exploration, or publication is as simple as sending a simple URL, which can be public or private.

O ADDRESS THE NEEDS of geographic data localization and strict data security requirements, SimpliFi has been rewritten to run inside Docker containers. Private SimpliFi[™] instances can be run on an on-premises server, or a virtual private cloud. Sensitive data won't ever leave the premises.

Quality Control & Visualization

NSURING QUALITY CONTROL IS IMPERATIVE in all analytical workflows. Particularly with extensive sample cohorts, mitigating batch effects and ensuring reproducibility becomes paramount. SimpliFi[™]'s latest features are designed to efficiently visualize extensive datasets and dynamically alert users to any anomalies (Fig. 1).



^{*}ProtiFi technologies are patent and patent-pending.

impliFiTM **APPLIES NONPARAMETRIC STATISTICS** to biological measurements, wherein sample replicates define their unique distributions. These distributions often deviate from normality, rendering many traditional statistical tools inadequate and prone to erroneous conclusions. Importantly, SimpliFi[™] always reports p-values and fold-changes with confidence intervals.

KEY FEATURE OF SimpliFi[™] IS ITS CAPABILITY to account for different data variances as a function of intensity of measurement: at low intensity, low stochastic sampling results in expectedly high variation, while at high intensity, effects like saturation become apparent. SimpliFi[™] accounts for these intensity-dependent variances.



HE IMPACT OF CONCORDANT OR DISCORDANT REPLICATES. As the number of biological replicates grows, consistency in observed changes between states leads to greater certainty in p-values. Conversely, inconsistencies or fewer replicates diminish certainty. Increasing the number of agreeing biological replicates enhances certainty, particularly when observed changes align. Conversely, discrepancies in observations or fewer replicates yield the opposite effect.

Note that **SimpliFi[™]** p-values represent the chance that control was chosen at random from experimental (or the inverse). This is not the same as the probability that two normally distributed populations have the same mean.



significant (p = 0.0006).





FFECT OF INTENSITY of observation: in comparing samples of the same kind (like control to control), one expects and sees that the ratios between any two samples is likely to be around 1, and that variability will exist (Fig. 2). The breadth of this curve, i.e. how often ratios between the same kind of sample will significantly deviate from 1, are strongly a function of intensity of observation.

- Interactive Plotting -

ATA CAN BE EXPLORED AND VISUALIZED with multiple interactive tools including volcano plots, distribution plots, heat maps, etc. SimpliFi™'s on-the-fly response yields a machine-human interface where human intuition guides data exploration guided. Users of all skill levels can take deep dives into the data and share projects via a simple URL.





NOTERSTANDING A DATASET requires understanding of regulation within the \mathbb{Y} biological systems. SimpliFiTM provides tools to map data to pathways and analyze cellular compartments and biological functions, as well as using the Reactome pathway database to quickly understand biological effects.

		Tumors (without outliers) - SimpliFi					
i	Centrol group Melanoma Synovial	Filters Hoorgoonetric p-value - < (0.05) (Text Gases In					
Ø	-	Natural killer T (NKT) cell	2.0545e-14 235 out of 4648	Mitotic fetal ger	m cell		Found
	Select Comparison	Uver bud hepatitic cell	2.0621e-5 20 out of 201	Show All Proteins			
í		Mitosic Initial gents cell	2.4233e-5	Accession	Gener same(s)		
	microsofe in results only analytics write	Leydy cell	1.6354e-3	P09670 P39117	VIM ETER FPSAS	Venentin Deckron transfer flavoprotein subunit beta	
(all		Contraction of Contra	26 out of 403	P31939	ATIC, OK/SW-cLB6, PURH	Bitunctional purine biosynthesis protein ATHC Extension for installant installant parties 3 adve	
9	in mushing a company bird Cool	Granulosa cell	21938e-3 15-out of 188	Q13895	BYSL ENP1	Bystin	10.7
	log 2 loid change - 0 up down +	Normalicial	2.3261e-3	007025	CIQBP, GCIQBP, HABPL SF2P32	Complement component 1 Q subcomponent	binding protein, m
¢\$7	Constant of the second se		4 out of 17	096538	TSPANIS, UN03042/PR09858	Tetraparin 18	Neochonona
land i	s 25 % missing values	Cornea	5.4736e-3	Q96K37-1	RTF3L4	Transcription factor BTF3 homolog 4	
	and the second se		3 out of 11	043828	RNU3P2, U355K, RRP9	U3 small nucleolar RNA-interacting protein 2	
5	sensitivity 2 0 %	Mesenchymal progenitor cell	6.239e-3	P36578	RPL1 RPL4	Large ribosomal subunit profein ul.4	
		DC1 K1+ receptor call	710860.3	P11216	PIGB	Glycogen phosphorylaxe, brain form	
	specificity≥ 0 %	Press program con	9 out of 99	P49406-	KIAA0004, MRPL15, MRPL19	Large ribosomal subanit protein 66.19m	
		interstitul cell	0.0136	P04060	CST6, STF8, CST8	Custatin-8	
	(Test aniech-		3 out of 15	P04350	TUBBA, TUBBS, TUBBAA	Tutxin beta-4A chain	
		konocyte cell	0.0144	Q92979 P67936	CDF, EMGI TPM4	Ribosomai 874A small subunit methyltransfer Tropomyosin alpha-4 chain	ate NDP3
	Only include favorite analytes		15 out of 296	000273-1	DFF1 DFF45, DFFA, H13	DNA fragmentation factor subunit alpha	
	Found 1074 of 5410 analytes.	Leydig precursor cell	0.0204	060220	DOP DOP1, TIMBA, TIMMBA, DEE2, EE2	Mitochondrial import inner membrane transic Elemention factor 2	Acase subunit Tenil
		Manager Trans	0.0222	Q9H3U5-2	SMAPL UNC45A	Protein unc-45 homailog A	
	Sort byp-value v	Wernory I cell	5 out of 47	P06733-1	ENGLENGEL MEPEL MPEL	Alpha-enolase	
	Sort orderNormal 👻	Fibrocartilage chondrocyte	0.0252	P23542-4	FILM, PP213	Fibulto-1	
	Fold change format log2 v	1. N. S	9 out of 122	P14618-2	CIP3, PKM	Pyruvate kinase PKM	
		Trophectodermi cell	0.0277	Q13308-1(-2)	CCK4, PTK7	Guarrine nucleofide-binding profess-like 3 Inactive tyrosine-profess kinase 7	
	Antibutor Dationana		9 out of 124	060504-2	SCAML SORES3	Winesin	
	An offer	Adjocyte	0.0342	Q7Z2W4 Q9H4A4	PROS577, 2C3HAV1, 2C3HDC2 APR, ENPEP	Zinc finger CCCH-type antiviral protein 1 Aminoposticlase B	
	GO Turnes Cell Markers	-	0.0200	P18077	GIG33 RPL35A	Large ribosomal subunit protein et.33	
	Contraction of the second second	PROCR+ progenitor cell	0.0386 3 put of 22	P62851	RPS25	Small ribosomal subunit protein e525 Series (Treasting contain kings) 10	
	Walcana Plat Viena Diagram	Karatriocota	0.041	Q99584	SIOCALS	Protein 5200-ALX	
	Contraction of the second	and a second sec	4 out of 38	Q9NRX2 REX164	LIP2, MRPL17	Large ribosomal subunit protein bL17m	
		Basal cell	0.0418	Q9H6R4-1(+2)	NOLS	Nucleolar protein 6	
			\$2 out of \$98	P62277	RP513	Small ribosomal suburit protein uS35	
		Bladder	0.0429	P62249	RPS16	Small ribosomal subunit protein uS9	
		-	13 0(10) 221	Q9Y5P6	CMPPE	Mannose-1-phosphate guarytransferase bet	
		Tendon	0.0433	297296-(*3) P62424	RFL7A, SURF-1, SURF3	Large ribosomal subunit protein eLB	1
		Missioners and	0.0433	054561	NDUFABL	Acyl carrier protein, mitochondrial	
			3 out of 23	P40925 (*1) P63247	MDHL MDHA FTEL MFTL RPSTA	Malate dehydrogenase, cytoplawnic Sonall rifecontal suburit protein eS3	
		Lymph	0.0495	P46783	RPS10	Small ribosomal subunit protein eS30	
			9 out of 138				



PATHWAY AND CELLULAR COMPARTMENT MAPS IN CONJUNCTION WITH GO ANNOTATION ALLOW FOR DEEP EXPLORATION OF PROTEOMICS, METABOLOMICS, LIPIDOMICS, TRANSKRIPTOMICS, AND GENOMICS DATASETS.

Conflict of Interest: The authors are the developers, inventors, and/or owners in or of ProtiFi LLC. Notwithstanding, we present these results as scientists.

-5-27

Data-to-Meaning via Reactome Integration

